第38卷 第5期
2021年9月

干 旱 区 研 究
ARID ZONE RESEARCH

Vol. 38 No. 5
Sep. 2021

基于EEMD-LSTM模型的天山北坡经济带年降水量预测

杨 倩[1,2]， 秦 莉[2]， 高 培[1,2]， 张瑞波[2]

（1. 中国气象局气象干部培训学院新疆分院，新疆 乌鲁木齐 830013；
2. 中国气象局乌鲁木齐沙漠气象研究所，新疆 乌鲁木齐 830002）

摘 要：降水量预测是现代气候预测业务的核心和难点。耦合模型在新疆降水量预测的研究应用屈指可数，因此，通过尝试建立集合经验模态分解（Ensemble Empirical Mode Decomposition，EEMD）和长短期记忆神经网络（Long Short-Term Memory Network，LSTM）的耦合模型对天山北坡经济带降水量进行预测研究。将1965—2019年天山北坡经济带共55 a的年降水量数据进行EEMD分解，转换成4个平稳分量和趋势项，通过谱分析得出各个分量的准周期，为后续训练LSTM模型提供基础。根据EEMD分解后的各分量训练得出LSTM网络模型并利用该模型进行研究区降水量预测。结果表明：EEMD-LSTM耦合模型预测2010—2019年天山北坡经济带降水量的平均相对误差为13.38%，均方根误差为38.03 mm，认为EEMD-LSTM耦合模型对天山北坡经济带降水量预测精度较好。利用EEMD-LSTM耦合模型预测2020—2029年天山北坡经济带年降水量，其中有6 a降水偏多，4 a降水偏少，2025年可能为极端湿润年，降水偏多超过20%；而2021年为极端干旱年，降水量预计低于200 mm。本文探索了干旱区降水量预测的新方法，并为气象防灾减灾工作提供参考依据。

关键词：集合经验模态分解；长短期记忆网络；年降水量；预测

气候预测是气象业务的核心与重点[1]，其中降水量预测由于降水事件的随机性和不确定性，难度较大。目前降水量预测的方法主要有统计预测和数值模式预测，对于年际和年代际降水量预测的主流方法仍然是统计学方法[2]。当前国内学者在年降水量的预测领域已经积累了大量的研究成果，传统的统计学方法如：自回归滑动平均（Auto-Regressive Moving Average，ARMA）模型[3]、小波分析[4-5]等已经不能满足降水量预测业务的需求，随着信号技术和机器学习的发展，年降水量的预测研究集中在了集合经验模态分解（Ensemble Empirical Mode Decomposition，EEMD）以及人工神经网络上[6-11]。不管是传统方法还是新技术，都是利用单一模型进行降水量的预测，由于单一模型的局限性，预测的准确率和精度都存在一定缺陷，预测效果不理想。为此将EEMD和人工神经网络相结合的预测方法，能有效的弥补单一模型的不足，并且非常适合降水量这一非平稳、非线性气象要素的预测[12]：其中EEMD分解可以将非平稳时间序列转换为平稳的本征模态函数分量；人工神经网络方法又能较好的拟合非线性关系。

天山是世界干旱区独具特色的山系，其拦截了西风气流携带的水汽，形成了较为丰沛的山区水资源[13]，由于天山山区复杂的地形地貌特征，气象站点稀少且分布不均给降水量预测研究带来了极大的困难。天山北坡经济带位于天山北麓，是新疆经济最发达、资源最优越的地区，同时作为丝绸之路经济带的重要枢纽[14]，其水资源分配不均、旱涝灾害频发等问题制约着当地的发展[15]，因此，提升天山北坡经济带降水量预测的准确率，对该区域的生产生活具有重要的应用价值。

目前，基于EEMD和人工神经网络的耦合模型

收稿日期：2020-10-14； **修订日期：**2021-01-09
基金项目：中国沙漠气象科学研究基金（Sqj2013005）；国家自然科学基金（41805130，41975110）；新疆维吾尔自治区重点实验室开放课题（2019D04002）；新疆维吾尔自治区天山青年计划（2020Q026）；新疆维吾尔自治区天山雪松计划（2019XS12）
作者简介：杨倩（1982-），男，高级工程师，主要从事气候变化研究. E-mail: 94306655@qq.com
通讯作者：秦莉. E-mail: qinhappy@sina.com

对新疆年降水量的预测研究还尚未展开,本文根据获取的研究资料、计算机响应速度等因素,选取了人工神经网络模型中更适于时间序列预测的长短期记忆(Long Short-Term Memory,LSTM)网络模型,并参考邸浩[16]构建的预测商品价格的EEMD-LSTM模型,建立了新疆天山北坡经济带EEMD-LSTM年降水量耦合模型,对该区域2020—2029年年降水量进行预测,并对模型进行评估,旨在为改进区域降水量预测模型提供新的参考,同时也为该区域的水资源开发利用、气象防灾减灾等提供决策依据。

1 数据与方法

1.1 数据来源与处理

本文选取了天山北坡经济带中16个气象站点(图1),1965—2019年共55 a的逐月降水量数据,通过对缺失值和异常值进行插补、剔除等处理后作为基础数据资料,再将各站资料平均后的年降水量序列数据进行EEMD分解,分析各分量信息特点,最后预测2020—2029年的年降水量。气象数据来源于新疆气象局气象信息中心,高程数据来源于通用大洋水深制图(General Bathymetric Chart of the Oceans,GEBCO)网站(https://www.gebco.net/)的全球海洋和陆地地形模型网格数据集(GEBCO_2020 Grid)。

1.2 研究方法

1.2.1 集合经验模态分解 经验模态分解EMD(Empirical Mode Decomposition)是一种处理非平稳时间序列的信号学方法,而集合经验模态分解(EEMD)是在经验模态分解的基础上通过插入高斯白噪声序列,有效抑制了EMD方法中出现的边缘效应和尺度混合现象[17],使最终分解的本征模态函数IMF(Intrinsic Mode Function)分量保持了物理上的唯一性,由于该方法无需预先设定任何基函数,同时添加的白噪声幅值对最终结果的影响较小,因此该方法具有较好的适应性[18]。降水量的变化是气候变化的具体体现之一,通过对降水量序列进行EEMD分解得到的各IMF分量隐藏在降水量序列中的不同时间尺度的气候振荡信号,而趋势项分量则反映了降水量的变化趋势情况。分析降水量各IMF分量序列的准周期信息,可以获取降水量变化的内在规律,为降水量的预测工作提供基础。

EEMD基本步骤如下:

首先,给原始信号 $x(t)$ 加入一组新的白噪声序列 $\omega(t)$,得到新信号序列为 $X(t)$,如公式(1):

$$X(t) = x(t) + \omega(t) \tag{1}$$

其次,对 $X(t)$ 进行EMD分解,得到一组IMF分量,如公式(2):

$$X(t) = \sum_{j=1}^{n-1} IMF_j(t) + r(t) \tag{2}$$

式中: $IMF_j(t)$ 表示第 j 个IMF分量; $r(t)$ 为趋势项。

然后对原始信号重复前两个步骤m次,每次都加入不同的白噪声,得到一组新的IMF分量,如公式(3)所示:

$$X_i(t) = \sum_{j=1}^{n-1} IMF_{ji}(t) + r_i(t) \tag{3}$$

将每次得到的IMF分量进行集成均值,使加入的白噪声互相抵消,作为最终结果。

经过上述的4个步骤,可得到各个尺度上的本征模态函数。

1.2.2 LSTM网络 LSTM网络是循环神经网络的一种变形结构,能够很好地挖掘数据信息传递中的长期依赖关系[19],可以有效缓解循环神经网络训练时的梯度消失问题。LSTM网络的单元结构如图2所

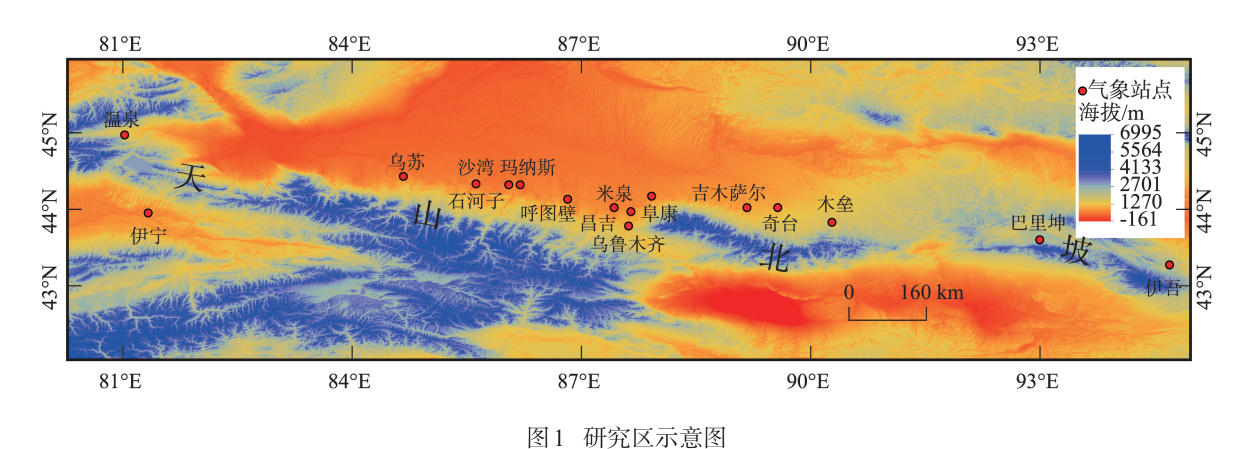


图1 研究区示意图

Fig. 1 Schematic diagram of research area

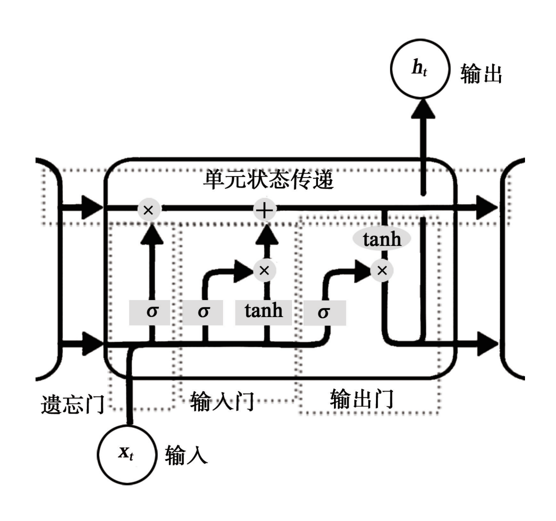


图2 LSTM网络单元结构

Fig. 2 Unit structure of LSTM neural network

示,包括遗忘门、输入门、输出门以及单元状态信息传递部分,左右两侧是时间顺序中相同结构的循环单元[20]。图2中的σ、tanh分别表示以sigmoid和双正切函数,为激活函数的非线性神经元。这些神经元的组合和运算(单元中圆圈部分)实现了门控制功能,并将运算结果传递给下一个单元结构。从下往上看,输入与输出之间存在非线性激活函数,确保神经网络能够逼近任意非线性函数[21];从左往右看,每个时刻的输入都能留存在网络中且对下一时刻产生影响,完成长短期记忆的传递。

1.2.3 评价指标 选用观测值与预测值之间的均方根误差RMSE、平均相对误差MRE、预测精确度P以及距平同号率作为模型的评价指标[22-23],计算公式如下所示:

$$\mathrm{RMSE}=\sqrt{\frac{1}{m}\sum_{t=1}^{m}(y_t-\hat{y}_t)^2} \tag{4}$$

$$\mathrm{MRE}=\frac{1}{m}\sum_{t=1}^{m}\frac{|y_t-\hat{y}_t|}{|y_t|} \tag{5}$$

$$P=\left(1-\frac{\frac{1}{m}\sum_{t=1}^{m}|y_t-\hat{y}_t|}{\frac{1}{m}\sum_{t=1}^{m}y_t}\right)\times 100\% \tag{6}$$

公式(4)~公式(6)中:y_t为t时刻的观测值;$\hat{y}_t$为t时刻的预测值。

2 结果与分析

2.1 EEMD-LSTM耦合模型的构建思路

论文采用EEMD-LSTM耦合模型进行天山北坡经济带年降水量的预测。具体的建模步骤包括:(1)对天山北坡经济带年降水量数据进行EEMD分解。(2)将EEMD分解后的各分量序列进行数据集的划分。将1965—1999年的分量数据作为训练集,2000—2009年的分量数据作为验证集,2010—2019年的分量数据作为测试集。(3)将各分量数据进行离差标准化处理,得到归一化数据。(4)进行LSTM网络的训练,确定网络的结构、参数与超参数。利用训练集数据训练LSTM网络,该网络结构由输入层、隐藏层和输出层的结构共同决定,其中输入层和输出层的结构由特征维度和序列的时间步长决定。LSTM网络的基础参数由监督式学习训练生成,而隐藏层神经元数量、迭代次数,Dropout等超参数通过试凑实验确定。(5)利用测试集数据评估EEMD-LSTM耦合模型的性能。利用训练好的模型预测研究区2010—2019年的降水量,通过与原数据对比,评估天山北坡经济带年降水量EEMD-LSTM耦合模型的优劣。(6)利用EEMD-LSTM耦合模型进行天山北坡经济带2020—2029年降水量的预测。

2.2 天山北坡经济带降水量的预测

2.2.1 天山北坡经济带降水量EEMD分解 对1965—2019年天山北坡经济带年均降水量序列进行EEMD分解,得到了4个IMF分量和1个趋势变化分量RES(图3)。

EEMD分解后的各分量包含了原序列中不同时间尺度的信息,同一时间段各分量信息的总和为原有序列的相应表征,各分量所含实际物理意义的多少可以通过显著性检验判断[17]。从1965—2019年天山北坡经济带年降水量EEMD分解后各分量的显著性检验结果可知(图4)。IMF_1分量在90%置信线上,包含的信息最为显著,而其他分量均在50%的置信线以下,认为其他分量包含的信息相对较少,但其仍然参与了方差贡献率的计算,保持了信号的总能量不变。

根据谱分析计算各分量的准周期,而各分量的频率与振幅对原始数据的影响程度可以用方差贡献率表示[17],由图3和表1可知,准周期为2.9 a的IMF_1分量方差贡献率为54.54%,振荡信号非常显著;IMF_2准周期为7.5 a,方差贡献率为15.70%,分别在1975—1985年和2003—2013年出现了两个相对平稳的时期;IMF_3分量的准周期为16 a,方差贡献率为7.75%,分量振幅随着时间推移呈现逐渐减小的

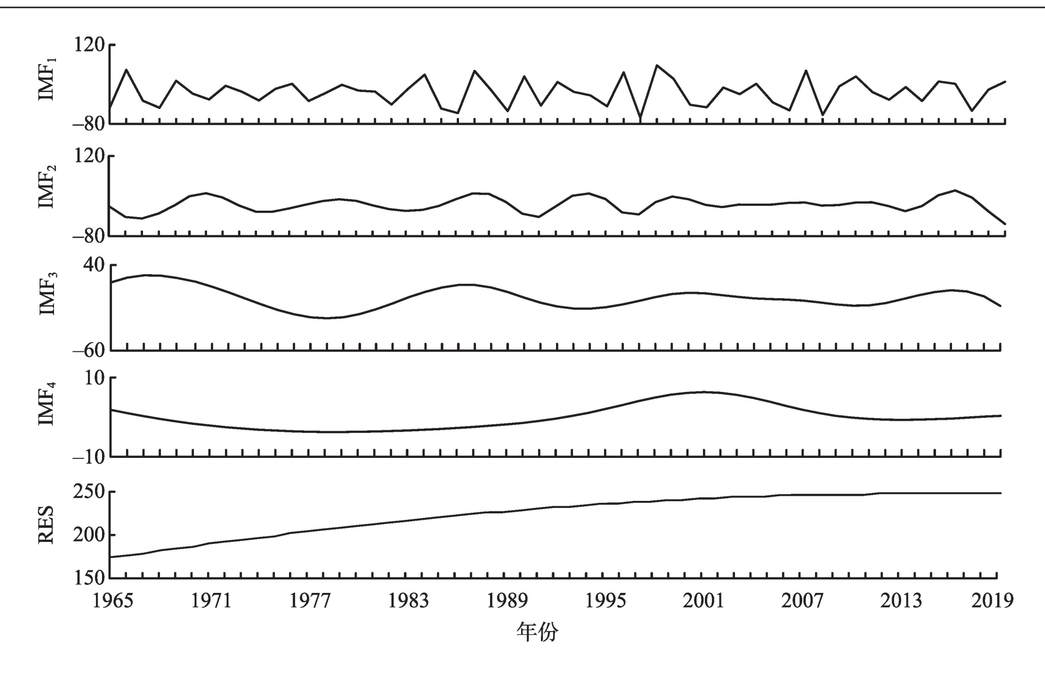
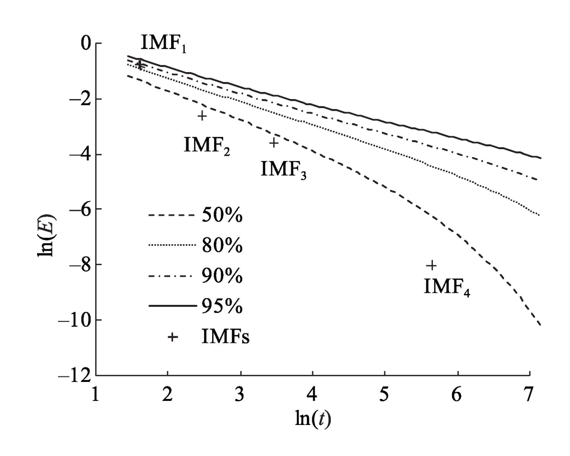


图3　1965—2019年天山北坡经济带降水量年际变化的EEMD分解

Fig. 3　EEMD decomposition of interannual variation of precipitation in the Northern Slope Economic Belt of Tianshan Mountains from 1965 to 2019



注:横坐标表示IMF分量对应周期的自然对数,纵坐标表示IMF分量的能量谱密度的自然对数。

图4　1965—2019年天山北坡经济带降水量各IMF分量的显著性检验

Fig. 4　Significance test of IMF components of precipitation in the Northern Slope Economic Belt of Tianshan Mountains from 1965 to 2019

表1　各分量准周期及方差贡献率

Tab. 1　Quasi period and variance contribution rate of each component

分量	准周期/a	方差贡献率/%
IMF_1	2.9	54.54
IMF_2	7.5	15.70
IMF_3	16.0	7.75
IMF_4	51.2	0.45

趋势;IMF_4振幅在10个单位左右变化,方差贡献率仅有0.45%;趋势项的方差贡献率为20.15%,研究区年降水量变化呈现出增速不断减缓的上升趋势。

新疆地区的水汽主要源自于地中海、北大西洋和北冰洋[24],在西风的输送下到达天山北坡,水汽输送过程主要受到北极涛动(Arctic Oscillation,AO)与北大西洋涛动(North Atlantic Oscillation,NAO)的影响[25],因此,天山北坡的降水准周期与AO、NAO的周期应具有一致性,而胡跃文[26]发现NAO与AO存在准8 a的周期,AO还存在准16 a的年代际周期。关学锋[27]也发现新疆北部地区的降水与北冰洋涛动指数(Arctic Oscillation Index,AOI)在准3 a周期的频域上对应很好,以上研究结果与本文的准周期结果相吻合。

2.2.2　LSTM网络模型训练　LSTM网络结构根据变量个数、输入层时间步长及预测的时间步长确定,网络模型的变量只有降水量,因此变量个数为1。输入层的时间步长以EEMD分解后各分量的周期作为依据,且时间步长应大于分量的周期小于原始数据的长度。根据此要求,参照IMF_3分量16 a的准周期,确定输入层的时间步长为17。模型的输出时间步长由预测年份的时间长度决定,故输出时间步长为10。

LSTM网络训练使用的激活函数为ReLU,网络

的优化函数为Adam函数，损失函数的目标为均方根误差。预防过拟合的方法采用Dropout方法，以提高LSTM网络的泛化能力。LSTM网络的超参数通过试凑实验确定。超参数实验的设计思路为固定两个超参数，试凑剩余的一个超参数，将训练出的LSTM网络预测值与验证数据集进行对比，选择均方根误差最小的值对应的超参数。最终LSTM网络的超参数如表2所示。

表2　各分量序列LSTM网络的超参数

Tab. 2　The hyperparameters of LSTM network with each component sequence

分量	隐藏层单元数/个	训练次数/次	Dropout
IMF_1	160	30000	0.5
IMF_2	100	600	0.5
IMF_3	16	5000	0.1
IMF_4	16	1000	0.2
RES	128	10000	0.5

注：Dropout表示神经网络单元被丢弃的概率值。

2.2.3　EEMD-LSTM耦合模型的测试与评估　利用构建的EEMD-LSTM耦合模型，预测天山北坡经济带2010—2019年年降水量，进而评估模型效果。图5为各分量序列的预测值与观测值的对比情况。IMF_1至IMF_4及趋势项的均方根误差分别为12.18 mm、28.82 mm、7.68 mm、1.85 mm、2.18 mm。

将各分量的预测结果求和，重建天山北坡经济带2010—2019年年降水量的预测值，并计算得出各年的绝对误差、相对误差和距平同号情况（表3）。从表3可以看出，天山北坡经济带2010—2019年，EEMD-LSTM耦合模型预测的相对误差最小为3.89%，出现在2013年，最大为30.79%，出现在2019年，前5 a平均相对误差为7.88%，后5 a的平均相对误差为18.87%，10 a的平均相对误差为13.38%，均方根误差为38.03 mm，距平同号率为60%。根据平均相对误差的结果及距平同号情况来看，该模型对超过5 a的降水量预测能力明显下降，但依据平均相对误差预测精度的分级标准[28]，平均相对误差介于

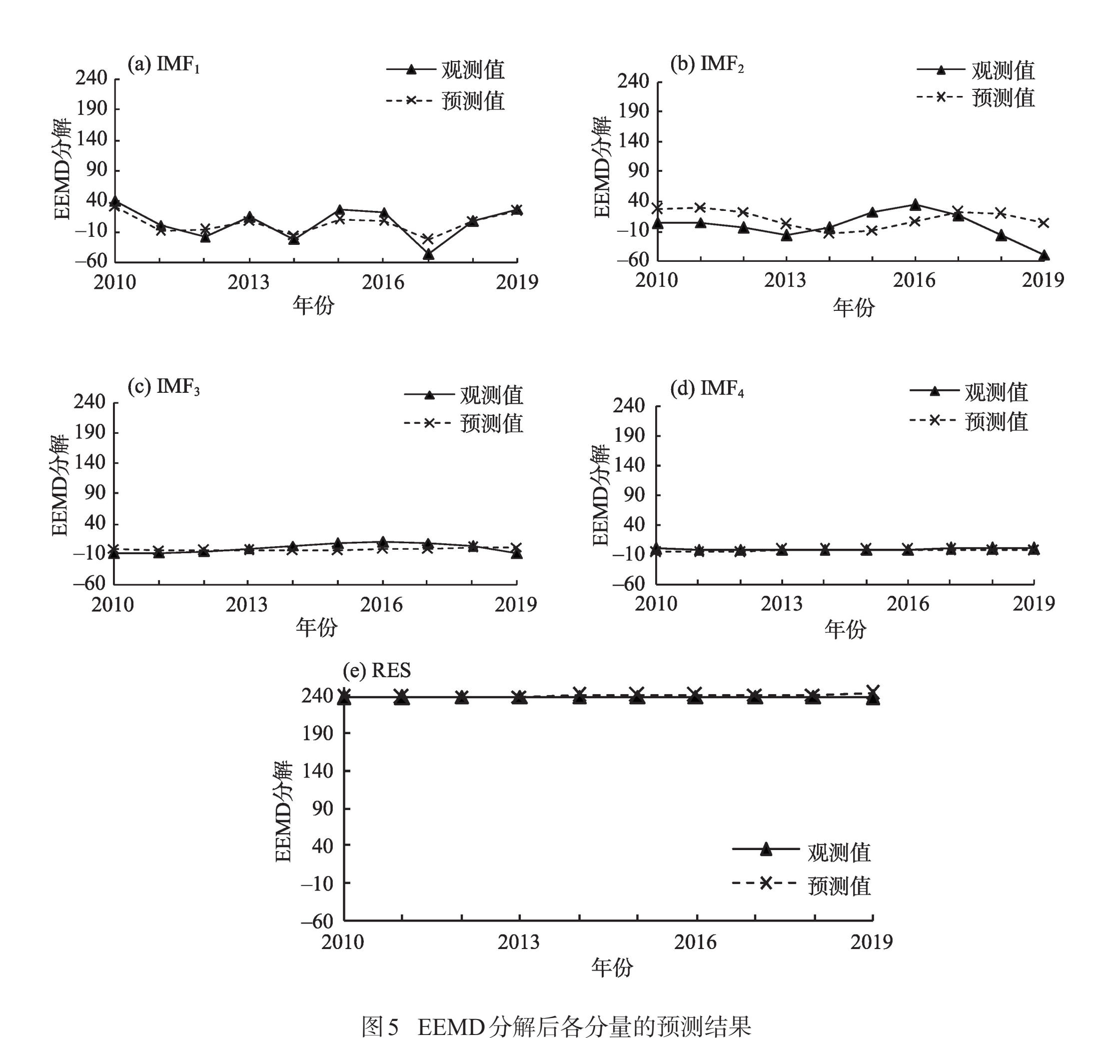


图5　EEMD分解后各分量的预测结果

Fig. 5　Prediction results of each component after EEMD decomposition

表3 2010—2019年天山北坡经济带年降水量LSTM网络预测结果

Tab. 3 LSTM network prediction results of annual precipitation in the Northern Slope Economic Belt of Tianshan Mountains from 2010 to 2019

	2010年	2011年	2012年	2013年	2014年	2015年	2016年	2017年	2018年	2019年
绝对误差/mm	16.44	17.48	37.22	9.07	9.38	59.48	52.35	20.17	39.3	64.04
相对误差/%	5.98	7.41	17.75	3.89	4.37	20.08	17.18	9.27	17.04	30.79
距平符号	同号	同号	异号	同号	同号	同号	同号	异号	异号	异号

10%~20%的预测效果较好，认为EEMD-LSTM耦合模型的整体预测效果较好。

由于LSTM网络的输入步长为17，模型预测的时间序列为1982—2019年共37 a的降水量数据。从图6可以看出，LSTM网络对天山北坡经济带1982—2019年年降水量的预测效果整体较好，平均绝对误差为22.17 mm，均方根误差为26.95 mm，平均相对误差为9.56%，距平同号率为81.58%，38 a中共有25 a的预测相对误差小于10%，集中出现在1982—2007年，而2015年、2016年和2019年的预测效果较差，主要原因由于EEMD-LSTM耦合模型对降水量突变的情况预测效果较差，而数据显示IMF_2分量在2016年的方差变化较大，从而增加了模型的预测误差。

为了进一步比较传统方法和EEMD-LSTM方法的优劣，采用自回归差分滑动平均(ARIMA)模型进行研究区降水量的预测，结果表明，ARIMA模型对1982—2019年年降水量预测的平均绝对误差为84.76 mm，均方根误差为102.97 mm，平均相对误差为38.69%，3种误差值均远远大于EEMD-LSTM耦合模型，即EEMD-LSTM耦合模型的预测效果优于传统模型。

利用配对样本T检验，验证模型的预测值与观测值是否存在显著差异。结果显示，两样本的相关系数为0.787，具有显著相关性，同时从配对样本T检验中可以看出(表4)，配对样本差值的平均值为3.24 mm，在95%的置信区间内，双尾检验的尾概率值P=0.47>0.05，故认为基于EEMD-LSTM耦合模型预测天山北坡经济带降水量的预测值和观测值之间不存在明显的差异。

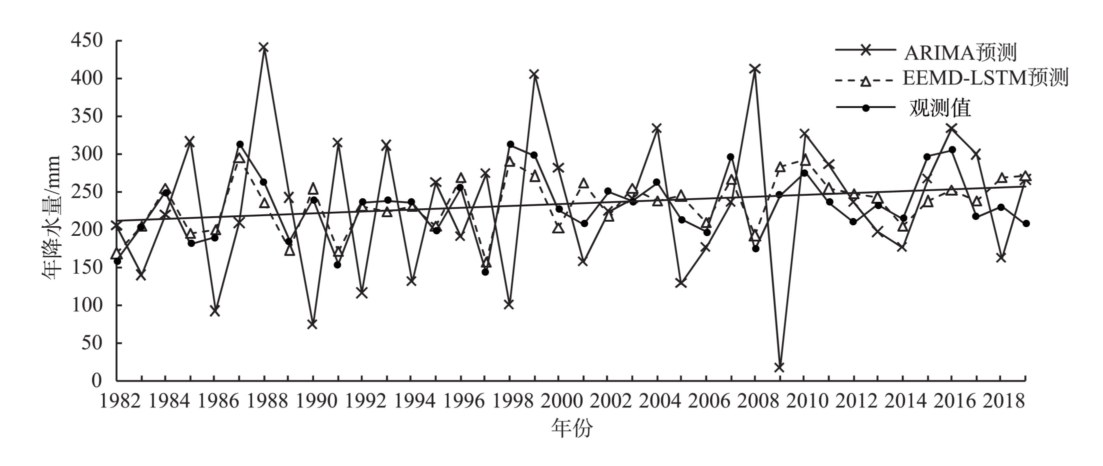


图6 1982—2019年天山北坡经济带年降水量的模型预测结果与实际数据对比

Fig. 6 Comparison of model prediction results and actual data of annual precipitation in the Northern Slope Economic Belt of Tianshan Mountains from 1982 to 2019

表4 LSTM网络预测值与观测值在测试集内的配对样本T检验

Tab. 4 Paired sample T-test table of LSTM network predicted value and observed value in test set

	配对差值					T	自由度	显著性(双尾)
	平均值	标准偏差	标准误差	差值的95%置信区间 下限	差值的95%置信区间 上限			
预测值与观测值	3.24	27.12	4.39	−5.68	12.15	0.74	37	0.47

2.2.4　天山北坡经济带年降水量预测　利用EEMD-LSTM耦合模型对2020—2029年天山北坡经济带年降水量进行预测(表5)，2020—2029年中有6 a降水偏多，4 a降水偏少，其中2025年的降水量为未来10 a的最大值，降水偏比多年平均偏多20%；而2021年为未来10 a降水量最小的年份，降水量低于200 mm。

表5　2020—2029年天山北坡经济带年降水量LSTM网络预测结果

Tab. 5　LSTM network forecast results of annual precipitation in the Northern Slope Economic Belt of Tianshan Mountains from 2020 to 2029

	2020年	2021年	2022年	2023年	2024年	2025年	2026年	2027年	2028年	2029年
降水量/mm	218.8	192.81	263.12	237.89	208.8	277.6	213.95	236.3	242.2	239.59
距平百分率/%	−5.38	−16.62	13.78	2.87	−9.71	20.04	7.48	2.18	4.74	3.61
丰枯情况	偏少	偏少	偏多	偏多	偏少	偏多	偏少	偏多	偏多	偏多

3　讨 论

天山北坡降水量的年际与年代际变化，主要受北半球中高纬度环流的影响，对AO、NAO等大尺度扰动有同频响应。有学者研究发现AO存在准3 a、准8 a、准16 a周期，NAO存在准8 a周期[25–27]。这些大气遥相关型与本文通过研究区年降水量EEMD分解所得的IMF分量在频率上基本一致，由此可以看出，EEMD能够自适应地提取出降水量信号中不同时间尺度的信息，分解后的分量具有一定的物理意义。同时EEMD也将非平稳的时间序列转变为多个平稳分量及趋势项之和，为后续建模奠定了良好的基础。

神经网络模型能够拟合任意的非线性关系，适用于降水量的建模预测。而神经网络模型泛化能力偏低，当数据形态变化较大时，长期预测精度将显著降低[29]。因此，平稳化的数据对神经网络模型预测精度的提升有着积极的意义，同时说明了EEMD与LSTM网络结合建模的必要性。本文建立的耦合模型对研究区降水量的预测效果较好，但仍有值得完善和提升的空间，下一步工作中利用再分析格点资料和大气环流因子作为模型的输入变量，丰富输入数据的类型。

4　结 论

本文建立了EEMD-LSTM耦合模型，对天山北坡经济带年降水量进行预测研究。首先通过对天山北坡经济带16个站点1965—2019年的降水量数据进行EEMD分解，将非平稳的降水量序列分解成了4个平稳的本征模态函数分量和1个趋势项，根据谱分析得出天山北坡经济带的降水量序列存在准2.9 a、准7.5 a、准16 a的周期变化。然后通过构建LSTM网络模型进行研究区年降水量预测，预测结果表明，该耦合模型的平均相对误差介于10%~20%之间，根据平均相对误差预测精度的分级标准，本文建立的EEMD-LSTM耦合模型预测效果较好，且优于传统的ARIMA模型，该研究为新疆降水量预测方法提供了新的思路和途径。

参考文献(References)：

[1] 刘颖, 任宏利, 张培群, 等. 中国夏季降水的组合统计降尺度模型预测研究[J]. 气候与环境研究, 2020, 25(2): 163−171. [Liu Ying, Ren Hongli, Zhang Peiqun, et al. Application of the hybrid statistical downscaling model in summer precipitation prediction in China [J]. Climatic and Environmental Research, 2020, 25(2): 163−171.]

[2] 任冉. 基于集合经验模态分解(EMD−EEMD)的中国夏季降水预报方法的研究[D]. 南京: 南京信息工程大学, 2014. [Ren Ran. The Research about the Summer Precipitation Prediction in China Based on EEMD Method[D]. Nanjing: Nanjing University of Information Science & Technology, 2014.]

[3] 迟道才, 张特男, 吴秀明, 等. ARMA模型在太子河流域年降水量预测中的应用[J]. 沈阳农业大学学报, 2012, 43(3): 607−610. [Chi Daocai, Zhang Tenan, Wu Xiuming, et al. Application of ARMA model in the annual precipitation forecast of Taizi River[J]. Journal of Shenyang Agriculture University, 2012, 43(3): 607−610.]

[4] 陈沪生, 周玉良, 周平, 等. 基于小波和ARIMA的黄山市年降水量分析及预测[J]. 南水北调与水利科技, 2019, 17(5): 50−55. [Chen Husheng, Zhou Yuliang, Zhou Ping, et al. Analysis and prediction of annual precipitation in Huangshan City based on wavelet and ARIMA[J]. South-to-North Water Transfers and Water Science & Technology, 2019, 17(5): 50−55.]

[5] 李佳秀, 杜春丽, 杜世飞, 等. 新疆极端降水事件的时空变化及趋势预测[J]. 干旱区研究, 2015, 32(6): 60−69 . [Li Jiaxiu, Du

Chunli, Du Shifei, et al . Temporal spatial variation and trend prediction of extreme precipitation events in Xinjiang[J]. Arid Zone Research, 2015, 32(6): 60−69.]

[6] 薛春芳, 侯威, 赵俊虎, 等. 集合经验模态分解在区域降水变化多尺度分析及气候变化响应研究中的应用[J]. 物理学报, 2013, 62(10): 1−7. [Xue Chunfang, Hou Wei, Zhao Junhu, et al. The application of ensemble empirical mode decomposition method in multiscale analysis of region precipitation and its response to the climate change[J]. Acta Physica Sinica, 2013, 62(10): 1−7.]

[7] 刘天虎, 刘天龙. 集合经验模态分解下中国新疆降水变化趋势的区域特征[J]. 沙漠与绿洲气象, 2015, 9(4): 17−24. [Liu Tianhu, Liu Tianlong. Regional features of precipitation variation trends over Xinjiang in China by the Ensemble Empirical Mode Decomposition method[J]. Desert and Oasis Meteorology, 2015, 9(4): 17−24.]

[8] 罗那那, 巴特尔·巴克, 吴燕锋. 基于集合经验模态分解北疆降水多尺度变化特征[J]. 水土保持研究, 2017, 24(4): 362−367. [Luo Nana, Bake Batur, Wu Yanfeng. Precipitation Multi-Scale characteristics by Ensemble Empirical Mode Decomposition in Northern Xinjiang[J]. Research of Soil and Water Conservation, 2017, 24(4): 362−367.]

[9] 党池恒, 张洪波, 陈克宇, 等. 长短期记忆神经网络在季节性融雪流域降水−径流模拟中的应用[J]. 华北水利水电大学学报(自然科学版), 2020, 41(5): 10−18, 33. [Dang Chiheng, Zhang Hongbo, Chen Keyu, et al. Application of the long short-term memory neural network for rainfall-runoff simulation in seasonal snowmelt basin[J]. Journal of North China University of Water Resources and Electric Power (Natural Science Edition), 2020, 41(5): 10−18, 33.]

[10] 吴麟, 冯利华. 基于BP神经网络的义乌市降水量预测[J]. 浙江水利科技, 2014, 42(2) : 52−54. [Wu Lin, Feng Lihua. Prediction of precipitation in Yiwu based on BP neural networks[J]. Zhejiang Hydrotechnics, 2014, 42(2): 52−54.]

[11] 沈艳, 杨春雷, 张庆国, 等. 基于RBF神经网络的池州市降水序列预测[J]. 安徽农业大学学报, 2012, 39(3): 451−455. [Shen Yan, Yang Chunlei, Zhang Qingguo, et al. Prediction of precipitation series based on RBF neural network in Chizhou City[J]. Journal of Anhui Agricultural University, 2012, 39(3): 451−455.]

[12] 张飞涟, 刘严萍. 经验模态分解与神经网络方法在降水预测领域的应用研究[C]//中国系统工程学会. 中国系统工程学会第十八届学术年会论文集—A01系统工程. 合肥: 中国系统工程学会, 2014. [Zhang Feilian, Liu Yanping. Application of Empirical Mode Decomposition and Neural Network in Precipitation Prediction[C]// Systems Engineering Society of China. Proceedings of the 18th Annual Meeting of Systems Engineering Society of China: A01 Systems Engineering. Hefei: Systems Engineering Society of China, 2014.]

[13] 宁理科. 地形地貌对天山山区降水的影响研究[D]. 石河子: 石河子大学, 2013. [Ning Like. Study on the Influence of Topography and Geomorphology on Precipitation over Tianshan Mountains, Central Asia[D]. Shihezi: Shihezi University, 2013.]

[14] 施雅风, 沈永平, 胡汝骥. 西北气候由暖干向暖湿转型的信号、影响和前景初步探讨[J]. 冰川冻土, 2002, 24(3): 219−226. [Shi Yafeng, Shen Yongping, Hu Ruji. Preliminary study on signal, impact and foreground of climatic shift from warm-dry to warm-humid in Northwest China[J]. Journal of Glaciology and Geocryology, 2002, 24(3): 219−226.]

[15] 方创琳. 天山北坡城市群可持续发展战略思路与空间布局[J]. 干旱区地理, 2019, 42(1): 1−11. [Fang Chuanglin. Strategic thinking and spatial layout for the sustainable development of urban agglomeration in northern slope of Tianshan Mountains[J]. Arid Land Geography, 2019, 42(1): 1−11.]

[16] 邸浩, 赵学军, 张自力. 基于EEMD-LSTM-Adaboost的商品价格预测[J]. 统计与决策, 2018, 34(13): 72−76. [Di Hao, Zhao Xuejun, Zhang Zili. Commodity price forecasting based on EEMD−LSTM-Adaboost[J]. Statistics & Decision, 2018, 34(13): 72−76.]

[17] Wu Z, Huang N E. Ensemble empirical mode decomposition: Anoise-assisted data-analysis method[J]. Advances in Adaptive Data Analysis, 2009, 1(1): 1−41.

[18] 郑一, 孙晓峰, 陈健, 等. 基于集合经验模态的随钻脉冲信号优良降噪算法[J]. 石油勘探与开发, 2012, 39(6): 113−116. [Zheng Yi, Sun Xiaofeng, Chen Jian, et al. Extracting pulse signals in measurement while drilling using optimum denoising methods based on the ensemble empirical mode decomposition[J]. Petroleum Exploration and Development, 2012, 39(6): 113−116.]

[19] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735−1780.

[20] 赵驰. 深度循环神经网络在特定场景自动问答中的应用研究[D]. 武汉: 武汉理工大学, 2018. [Zhao Chi. Application Research on Deep Recurrent Neural Networks in Automatic Question-Answer under Specific Occasions[D]. Wuhan: Wuhan University of Technology, 2018.]

[21] Graves A. Long Short-Term Memory: Supervised Sequence Labelling with Recurrent Neural Networks[M]. Berlin, Heidelberg: Springer, 2012: 1735−1780.

[22] 朱灵子. 基于统计模型的汛期降水预测研究——以义乌市为例[D]. 金华: 浙江师范大学, 2015. [Zhu Lingzi. Study of Flood Season Precipitation Prediction Based on Statistical Model: Taking Yiwu City as an Example[D]. Jinhua: Zhejiang Normal University, 2015.]

[23] 白慧, 高辉, 刘长征, 等. MODES系统对贵州月气温、降水预测初步评估[J]. 沙漠与绿洲气象, 2016, 10(5): 58−63. [Bai Hui, Gao Hui, Liu Changzheng, et al. Assessment of multi-model downscaling ensemble prediction system for monthly temperature and precipitation prediction in Guizhou[J]. Desert and Oasis Meteorology, 2016, 10(5): 58−63.]

[24] 杨莲梅, 刘晶. 新疆水汽研究若干进展[J]. 自然灾害学报, 2018, 27(2): 1−13. [Yang Lianmei, Liu Jing. Some advances of water vapor research in Xinjiang[J]. Journal of Natural Disasters, 2018, 27

(2): 1-13.]
[25] 戴新刚, 汪萍, 张凯静. 近60年新疆降水趋势与波动机制分析[J]. 物理学报, 2013, 62(12): 527-537. [Dai Xingang, Wang Ping, Zhang Kaijing. A study on precipitation trend and fluctuation mechanism in northwestern China over the past 60 years[J]. Acta Physica Sinica, 2013, 62(12): 527-537.]
[26] 胡跃文, 杨小怡. 北极涛动与北大西洋涛动的低频变化特征[J]. 气象科学, 2007, 99(3): 84-90. [Hu Yuewen, Yang Xiaoyi. Low-frequency variability of AO and NAO[J]. Journal of the Meteorological Sciences, 2007, 99(3): 84-90.]
[27] 关学锋, 孙卫国, 李敏姣, 等. 1965—2012年新疆北部地区气候变化及其对北极涛动的响应[J]. 干旱区研究, 2016, 33(4): 681-689. [Guan Xuefeng, Sun Weiguo, Li Minjiao, et al. Climate change in north Xinjiang and its response to Arctic Oscillation during the period of 1965-2012[J]. Arid Zone Research, 2016, 33(4): 681-689.]
[28] 刘德钊, 周海东, 荣莉, 等. GNNM(1, 1)模型在城市用水量预测中的应用[J]. 山东建筑大学学报, 2006, 21(4): 335-337. [Liu Dezhao, Zhou Haidong, Rong Li, et al. Predicting water demands of cities by GNNM (1, 1)[J]. Journal of Shandong Jianzhu University, 2006, 21(4): 335-337.]
[29] 王婷婷, 钱晓东. 时间序列的非线性趋势预测及应用综述[J]. 计算机工程与设计, 2010, 31(7): 1545-1549. [Wang Tingting, Qian Xiaodong. Survey of non-linear trend forecast and application of time series[J]. Computer Engineering and Design, 2010, 31(7): 1545-1549.]

Prediction of annual precipitation in the Northern Slope Economic Belt of Tianshan Mountains based on a EEMD-LSTM model

YANG Qian[1,2], QIN Li[2], GAO Pei[1,2], ZHANG Ruibo[2]
(1. Xinjiang Branch China Meteorological Administration Training Centre, Urumqi 830013, Xinjiang, China;
2. Institute of Desert Meteorology, CMA, Urumqi 830002, Xinjiang, China)

Abstract: Precipitation prediction is both an essential and challenging component of modern climate prediction. In the precipitation prediction in Xinjiang, the research and application of coupling models are very limited. Therefore, this paper attempts to establish the ensemble empirical mode decomposition (EEMD) and long short-term memory (LSTM) neural network coupling models to predict precipitation in the Northern Slope Economic Belt of Tianshan Mountains. Firstly, the precipitation data recorded in the study area during 55 years, from 1965 to 2019, were decomposed into four stationary components and trend terms using the EEMD, and the quasi period of each component was obtained by spectral analysis, which provided the basis for the subsequent training of the LSTM model. Then, each EEMD component was trained into the LSTM network model, and the models were used for predictions. The reconstruction and comparison of the results revealed that the average relative error and root mean square error of the 2010-2019 model were 13.38% and 38.03 mm, respectively. Therefore, it is inferred that the EEMD-LSTM coupling model can achieve a better precipitation prediction accuracy in the study area. The model was used to predict annual precipitation in the Northern Slope Economic Belt of Tianshan Mountains from 2020 to 2029; within this period, six years presented more and four years presented less precipitation. Year 2025 is predicted to be an extremely humid year with more than 20% of precipitation in excess; while 2021 is anticipated as an extremely dry year, with expected precipitation amounting to less than 200 mm. This study explored a new method of precipitation prediction in an arid area, and provided a reference for meteorological disaster prevention and mitigation.

Keywords: ensemble empirical mode decomposition; long short-term memory network; annual precipitation; prediction